

**IMAGE BASED INQUIRY SYSTEM FOR SEARCH
ENGINES FOR MOBILE TELEPHONES WITH
INTEGRATED CAMERA**

**Hartmut Neven, senior
Wilhelm Ziemonstr. 11
52078 Aachen
Deutschland**

Abstract

Mobiltelefone und Computer werden zunehmend mit einer Kamera ausgestattet. Dies ermöglicht es, statt bloßer Texteingaben auch Bilder an Suchmaschinen oder Datenbanken als Eingabe zu senden. Fortschritte bei Bilderkennungsverfahren wiederum ermöglichen zunehmend das automatische Erkennen von Objekten, Buchstabenfolgen oder Symbolen in digitalen Bildern. Dies erlaubt es, die Bildinformation in ein symbolisches Format, z. B. Klartext umzuwandeln, um damit Informationen zu dem gezeigten Objekt abzurufen.

Beschreibung

Ein Mensch sieht einen Gegenstand und sofort stellt sein Gedächtnis Informationen bereit, die mit dem Gegenstand in Zusammenhang stehen. Extrem nützliche wäre ein System, dass diese Leistung nachbildet oder sogar erweitert.

Moderne Verfahren der Bilderkennung erlauben es zunehmend besser Objekte, Landschaften, Gesichter, Symbole, Buchstabenfolgen, etc. in Bildern zu erkennen. Mehr und mehr Kameras sind an Geräte angeschlossen, die an Datenfernübertragungsnetzwerke angebunden sind. Solch eine Konfiguration unterstützt die folgende Anwendung. Mit der Kamera in einem Endgerät (1), z. B. in einem Mobiltelefon wird ein Bild oder eine kurze Bildsequenz aufgenommen. Dieses Bild (2) oder diese Bilder werden dann per Datenfernübertragung (3) an einen Serverrechner (7) geschickt. Dort läuft ein Bilderkennungsverfahren (4), das die Bildinformation in symbolische Information (5), z. B. Klartext umwandelt. Z.B. erkennt das Bilderkennungsverfahren, dass auf dem Bild der Eifelturm zu erkennen ist. Alles weitere funktioniert nun ähnlich wie bei einer traditionellen Suchmaschine (6) im Internet. Der Serverrechner schickt dem Nutzer eine Liste zurück mit "Links" auf Datenbankinträge oder Webseiten, die Informationen über das gezeigte Objekt (8) enthalten.

1. Bilderkennung

Dieser Abschnitt gibt einen groben Ueberblick ueber eine moegliche Methode zur Objekterkennung. Eine genauere Beschreibung zu Verfahren fuer die Objecterkennung ist in den folgenden Publikation beschrieben: J. Buhmann, M. Lades and C. v.d.Malsburg, "Size and Distortion Invariant Object Recognition by Hierarchical Graph Matching," in *Proceedings of the IJCNN International Joint Conference on Neural Networks, San Diego 1990*, pp. II-411-416 und "High-Level Vision : Object Recognition and Visual Cognition", Shimon Ullman, *MIT Press; ISBN: 0262710072; July 31, 2000*. Verfahren zur automatischen Schriftzeichenerkennung sind beschrieben in: "Optical Character Recognition: An Illustrated Guide to the Frontier" *Kluwer International Series in Engineering and Computer Science, 502*, by Stephen V. Rice, George Nagy, Thomas A. Nartker, 1999.

1.1. Aufbau einer Objektrepraesentation

Die meisten Objekterkennungsverfahren die heute verwendet werden, benutzen eine Anzahl von Beispielbildern (21) um dem Objekt angepasste Merkmalsdetektoren (22) zu trainieren. Dabei

1.2. Erkennung

Bei der Erkennung werden die trainierten Merkmalsdetektoren (32) verwendet um die von ihnen repraesentierten Merkmale in einem Eingabebild (31) aufzufinden. Dieses geschieht durch einen Suchprozess. Jeder Merkmalsdetektor gibt einen Konfidenzwert aus, der angibt, wie gut er das von ihm repraesentierte Merkmal in dem Bild erkennt. Wenn die akkumulierten Konfidenzwerte (33) aller Merkmalsdetektoren einen vorgegebenen Schwellenwert ueberschreiten nimmt man an, dass das Objekt erkannt wurde.

2. Anwendungsbepiele

Natuerlich ist die automatische Bilderkennung noch weit davon entfernt die Leistungen des menschlichen Sehsystems zu erreichen. Daher wird man sich zunaechst auf Situationen beschraenken, die von existierenden Bildverarbeitungssystemen gut behandelt werden koennen. Im folgenden beschreibe ich eine Reihe von Anwendungsfeldern und beschreibe ihre spezifischen Schwierigkeiten.

Stadt- und Museumsfuehrer

Visuell Gebaeude zu erkennen ist mit heutigen Methoden gut realisierbar. Es hilft natuerlich wenn der Nutzer das Gebaeude frontal und senkrecht fotografiert und nicht aus einem schraegen Winkel. Des weiteren kann man die Bilderkennung unterstuetzen in dem man Positionsinformationen mitverwendet. Viele Telefone werden mit GPS (Global Positioning System) ausgestattet, so dass man jederzeit bis auf wenige Meter weiss, wo

sich das Telefon befindet. Diese Information kann man nutzen um bei der Bildverarbeitung nur solche Gebäude oder Gebäudedetails in Betracht zu ziehen, die in Nähe sind. Da das Gebäude zu verschiedenen Tageszeiten erkennbar sein soll, muss man beim Aufbauen der visuellen Representation darauf achten, das entsprechendes Bildmaterial mit aufgenommen werden muss. Fuer die meisten Bilderkennungsverfahren bedeutet das, das man einfach mehrere Bilder unter verschieden Beleuchtungssituationen aufnimmt und diese bei der Modellkonstruktion verwendet.

Sehr einfach waere es auch, einen universellen Kunstfuehrer zu bauen, der einem Informationen, z. B. zu einem Gemälde gibt. Da Bilder zweidimensional sind ist die Erkennung wesentlich vereinfacht.

Produktinformationen

Ein andere Kategorie von Objekten sind Produkte wie Autos, Bücher oder Spielzeuge. Sieht der Nutzer ein Automodell das ihn interessiert, kann er einfach ein davon Bild aufnehmen und er wird z. B. zu einer entsprechenden Webseite mit weiteren Produktinformationen geleitet. Wiederum wird es in den frühen Phasen eines solchen Services nützlich sein, wenn der Nutzer Fotos von exakten Frontal- oder Seitenansichten aufnimmt und zum Serverrechner schickt. In späteren Versionen, wenn die Poseninvarianz verbessert worden ist, braucht sich der Nutzer weniger einzuschraenken. Es ist wichtig, den bildbasierten Suchservice so zu gestalten, das es ähnlich wie beim jetzigen World Wide Web jedem Anbieter von Informationen ermöglicht wird für seine Webseite eine bildbasierte Suchfunktion anzubieten. Auf diese Weise kann leicht sichergestellt werden, dass für viele Produkte eine bildbasierte Suchfunktion zur Verfügung steht, da z.B. Autohersteller ein großes Interesse daran haben werden, das ihre neuesten Modelle per Bildaufnahme erkannt werden können.

Texterkennung

Ein weiterer nützlicher Service besteht darin, das man Texterkennung abietet. Fuer den Reisenden nach Tokyo oder Paris, der der Landessprache nicht mächtig ist, waere es von grossem Wert, wenn er seine Kamera auf ein Schild richten kann und er dann eine Uebersetzung und weitere Informationen zu dem erkennen Text erhält. Steht man beispielsweise in Tokyo vor ein Sushibar waere es doch von grossem Wert, wenn man sofort und muehelos den entsprechenden Eintrag in einem Restaurantfuehrer lesen könnte. Gerade fuer Besucher, die japanische Schriftzeichen nicht lesen können, ist dies eine sehr bequeme Loesung um an weitere Informationen zu kommen.

Gesichtserkennung

Gesichtserkennung ist ein weiterer Spezialfall. Menschen, die aus irgendwelchen Gründen möchten, dass man schnell mehr über sie erfahren kann, können Aufnahmen von ihrem Gesicht verfügbar machen, die dann von der Bilderkennung genutzt werden können.

Das vollausgebaute System

Die Zahl der Anwendungsbereiche liesse sich noch lange fortsetzen. Kataloge fuer Antiquitaeten, Planzen- und Tierbestimmungsbuecher koennen mit dem beschriebenen System wesentlich effizienter gestaltet werden. Oder man stelle sich einen Teil einer Apparatur vor, fuer das man Ersatz oder weitere Erklaerungen braucht. Man nimmt einfach ein Bild auf und schnell wird man auf Kennung und Hersteller oder einen entsprechenden Abschnitt in einem Handbuch verwiesen. Ein System, das einem Zusatzinformationen zu Reklametafeln gibt, ist eine weitere Anwendung. In jedem dieser Faelle nimmt der Nutzer einfach ein Bild des Gegenstandes von Interesse auf und schickt es zum Rechner auf dem die Bilderkennung laeuft. Die Bilderkennung sendet enstprechende symbolische Information, die das Objekt beschreibt an die Suchmaschine, die letztlich die Information, die zum Nutzer geschickt wird, auswaehlt.

In der vollen Ausbaustufe hat man ein System, das man mit einem externen visuellen Gedaechtnis vergleichen koennte. Jeder Gegenstand, jeder Text, jedes Symbol, jedes Gesicht, letztlich eine grosse Anzahl von Ansichten der Erdoberflaeche ist in dem System gespeichert und wird kontinuierlich durch die Nutzer auf dem neuesten Stand gehalten. Letzlich hat man ein globales System, das unser Wissen ueber die Dinge auf unserem Planeten speichert und jederzeit zur Verfuegung stellt.

Zeichnungen

Abbildung 1

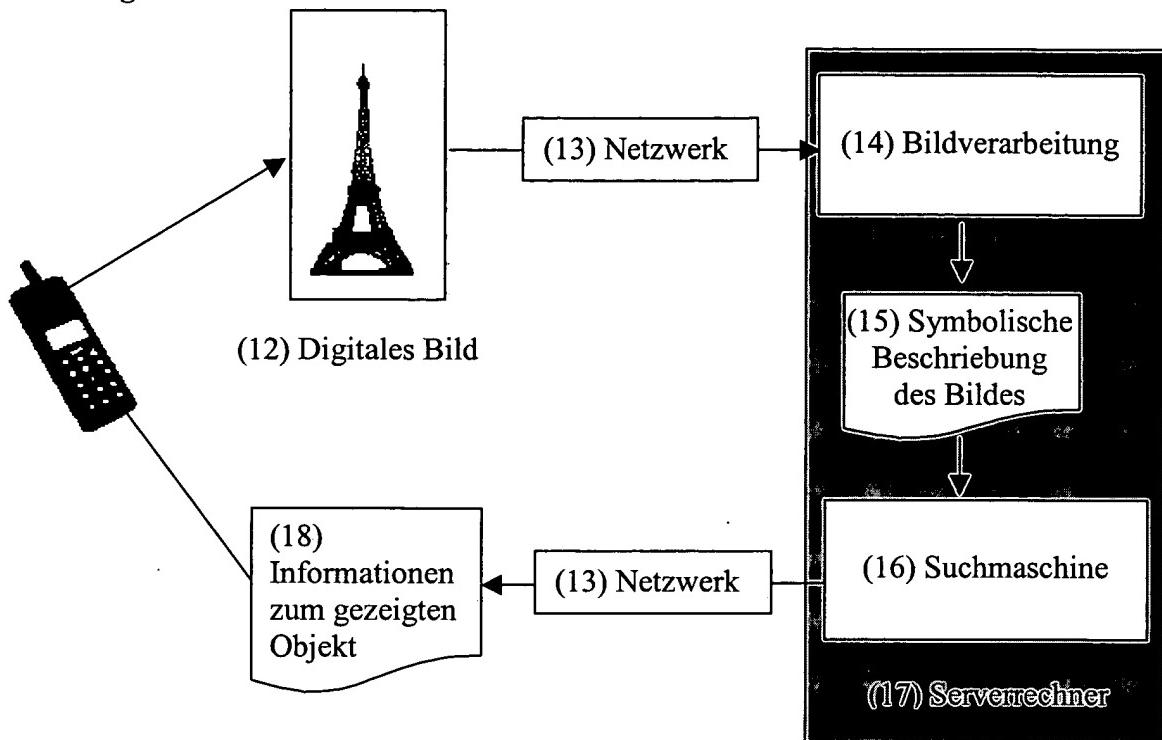


Abbildung 2

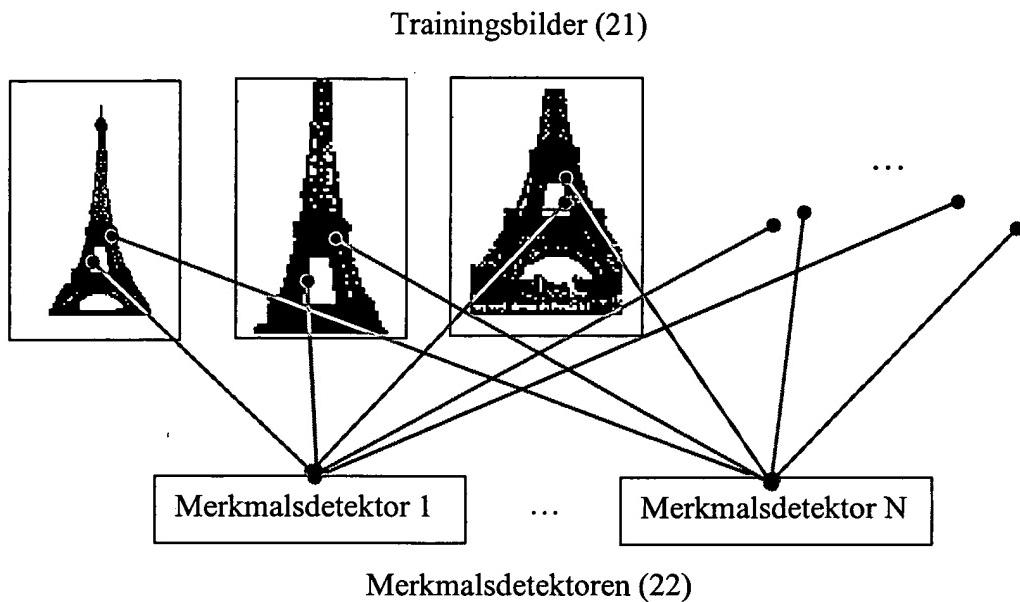
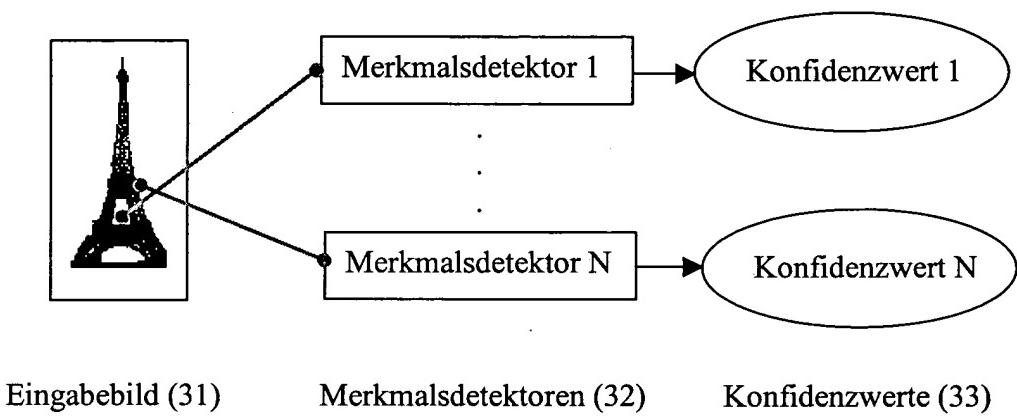


Abbildung 3



Ansprueche

- 1) Ein System zur bildbasierten Anfrage an Suchmaschinen oder Datenbanken, gekennzeichnet durch a) Ein Endgeraet mit eingebauter Kamera, welches an ein Datenfernuebertragungsnetz angeschlossen ist. b) Ein Serverrechner, auf dem ein Programm zur Objekterkennung laeuft, welches eingesandte Bilder analysiert und mit einer symbolischen Indizierung versieht. c) Eine Suchmaschine, welche die Bildindizes nutzt um Informationen zu dem Bild zu finden und zu dem Endgeraet zurueckzuschicken.
- 2) Ein System, wie beschrieben unter 1), das fuer Mobiltelefone oder mobile Computer ausgelegt ist, die eine eingebaute Kamera haben.
- 3) Ein Stadt- oder Musuemsfuehrer, der das unter 2) beschriebene System verwendet um einem Nutzer Informationen zu geben zu Objekten, von denen er zuvor ein Bild aufgenommen hat.
- 4) Ein System wie unter 3) bei dem zusaetzlich Positionsinformation verwendet wird um die Bilderkennung geeignet einzuschränken.
- 5) Ein System wie unter 2), das Produktinformationen bereitstellt zu Produkten die zuvor mit der mobilen Kamera fotografiert wurden.
- 6) Ein System wie unter 2), bei der die Objekterkennung auch in der Lage ist Textzeichen oder Symbole zu erkennen.
- 7) Ein System wie unter 2), bei der das System insbesondere in der Lage ist, Gesichter zu erkennen.
- 8) Ein System, wie unter 2), das genutzt wird, dem Nutzer zusaetzliche Information zu Reklametafeln zu geben.
- 9) Ein elektronisches Bedienungshandbuch, welches ein System wie unter 2) benutzt um schnell Zugang zu entsprechenden Abschnitten im Handbuch zu navigieren.
- 10) Ein System, wie unter 2), das es den Anbietern von Informationen ermöglicht, selbstständig neue Einträge in das Bildverarbeitungssystem vorzunehmen, um so die Abfrage ihrer Daten per Bildeingabe zu ermöglichen.